

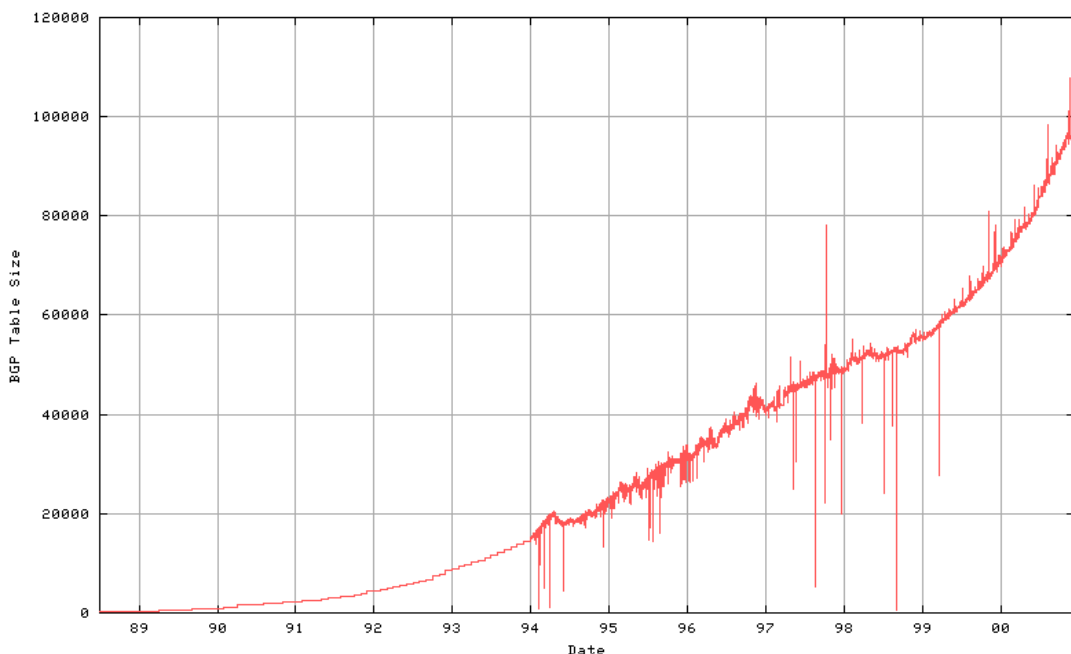
Scaling the Internet – The Routing View

Geoff Huston
February 2001

Last month, in looking at the structure of routing within the Internet I ended with the Big Question: how will routing deal with the demands of tomorrow's Internet? Lets take a quick look.

Graphs of the progress of almost any Internet-related metric have the same pattern of phenomenal growth. Doubling every year, or more, is the typical pattern. The number of domain names registered each year, the number of connected computers, the amount of traffic carried, the bandwidth of Internet backbone links. All these metrics show the same pattern of inexorable growth. And it seems like we are in for more. The current introduction of 3rd generation mobile wireless systems heralds a new wave of expansion for the Internet as we turn mobile telephones into mobile Internet devices.

The routing system is not immune to these pressures of growth. One of the best places to see this is in the size of the exterior routing table of the Internet. This table is the complete set of routes that describe the origin of every routed address within the Internet. As new networks connect to the Internet they announce their address prefix into this table. As the Internet grows so does the size of this table. Looking at this table at regular intervals can give us a good idea of what is happening within the routing system. This exercise started in 1988 using monthly samples of the size of the routing table. In 1994 Erik Jan Bos of Surfnets in the Netherlands started doing the same, but using hourly samples. The author continued this approach in 1997, and today we have a very detailed view of the dynamic of the BGP routing table for the past seven years, and a more general view that stretches back to 1988. Here's what it looks like.



There's quite a story behind this chart, and it can tell us a lot about what is likely to happen in the future. The chart appears to have four distinct phases: exponential growth between 1988 and 1994, a correction through 1994, linear growth from 1995 to 1998 and a resumption of exponential growth in the past two years.

Prior to 1994 the Internet used a routing system based on classes of addresses. One half of the address space was termed class A space, and used a routing element of 8 bits (or a /8) and the remaining 24 bits was used to number hosts within the network. One quarter of the space was termed class B space, with 16 bits of routing address (/16) and 16 bits of host address space, and one eighth was the Class C space, with 24 bits of routing address (/24) and 8 bits of host space. According to the routing system routed networks came in just three sizes, small (256 hosts), medium (65,535) hosts and large (16,777,215 hosts). Real routed networks however came in different sizes, most commonly one or two thousand hosts. For such networks a Class B routing address was a severe case of oversupply of addresses, and the most common technique was to use a number of Class C networks. As the network expanded so did the number of Class C network routes appearing in the routing table. By 1992 it was becoming evident that if we didn't do something quickly the routing table would expand beyond the capabilities of the routers of the day.

The solution was termed 'CIDR' or Classless Inter-Domain Routing. The technique was elegant and effective. Instead of dividing the network into just three fixed prefix lengths, you allow each routing advertisement to have an associated prefix length. If you use 4 Class C address blocks within your network, then, as long as the addresses were aligned correctly you can advertise them into the routing system using a single prefix of 22 bits. With some concerted effort in the operations community 1994 saw the widespread introduction of the BGP4 routing protocol and CIDR into the Internet's routing system. And the results were very effective. While the Internet doubled in size through 1994, the routing table remained pretty constant at some 20,000 routes.

CIDR led to the other change in routing policy, that of provider-based addresses and provider route aggregation. Instead of allocating network address blocks to every network, the address registry allocated a larger address block (a /19 prefix) to a provider, who in turn allocated smaller address blocks from this block to each customer. Now a large number of client networks would be encompassed by a single provider routing advertisement. This technique, hierarchical routing, is used in a number of network architectures, and is a powerful mechanism to aggregate routing information.

Through 1995 to 1998 the combination of CIDR and hierarchical provider routing proved very effective. While the Internet continued to double in size each year (or more!), the routing space grew at a linear rate, increasing in size by some 10,000 routes per year. For the routing system this was good news. Vendors were able to construct larger routers at a pace that readily matched the growth of the Internet, and a combination of Moore's law in hardware and CIDR and hierarchical routing in the routing system proved very effective in coping with dramatic growth in the Internet.

But midway through 1998 something changed. The routing system stopped growing at a linear rate and resumed a pattern of exponential growth again, at a rate of a little under 50% per year. This is a worrying pattern. While the size of the routing table is some 100,000 routes at the end of 2000, in a years time it will be some 150,000 routes and 225,000 routes the year after, and so on. Within six years the table will be reach some 1,000,000 routes at this rate of growth.

What is driving this recent change to exponential growth of the routing table?

In a word, *multi-homing*. Multi-homing is when an ISP has a number of external connections to other networks. This may take the form of using a number of upstream ISPs as service providers, or using a combination of upstream providers and peer relationships established either by direct links or via a peering exchange. The way in which multi-homing impacts the

global BGP table is that multi-homing entails pushing small address fragments into the global table with a distinct connection policy. What we are seeing in this sharp rise in the size of the BGP table is a rapid increase in the number of small address blocks being announced globally.

The driving factor behind such multi-homing is an effort to further reduce connectivity costs for the ISP and at the same time attempt to improve the resiliency of the service provided by the ISP to its customers. Using two or more upstream ISPs allows the ISP to switch its traffic from one to the other in the event of routing or connectivity failures in any single upstream ISP. It also allows the ISP to continually engineer its traffic flows between upstream providers in order to minimize the total costs of the upstream service. Peer relationships also help in a similar way. Traffic passed along a peering path incurs no cost to the ISP, unlike traffic passed to an upstream ISP, so that the cost of establishing a point of presence at a peering exchange can be offset by savings made in reduced traffic levels passed to upstream providers. Additional external connections also improve the resiliency of the overall service, establishing alternate paths to destinations that can be used in the event of failure of the primary path.

If this is as wonderful as it seems, then why didn't we think of this before, and why was the practice of single-homed provider-based hierarchies so common in the past? It looks like the determining factor is the cost of communications bearers. Connecting to multiple upstream services and connecting to peering exchanges implies the use of more access circuits. While the cost of these circuits was high, the offset in terms of benefit was low enough as to negate most of the potential benefits of the richer connectivity mesh. Over the past few years the increasing level of competition in the largely deregulated activity of provision of communications bearers has brought about reductions in the price of these services. This, coupled with an increasing technical capability in the ISP sector, has resulted in the increasing adoption of multi-homed ISPs. Of course it's not just ISPs. Many customers now are absolutely reliant on the Internet to operate their own business, and in the quest for ever increasing resiliency we are also starting to see multi-homed customers.

The effects on this will not just be felt with the rapid growth of the BGP table. If multi-homing becomes a common option of corporate customers, then what is happening is that the function of providing resiliency within a network has shifted from a value-added role within the network to that of a customer responsibility. And if customers are not prepared to pay for highly robust network services from any single ISP then there is little economic incentive for any single ISP to spend the additional money to engineer robustness within their service. From that perspective, what the ISP industry appears to be heading into is a case of a somewhat disturbing self-fulfilling prophecy of minimalist network engineering with no margin for error. But then, as the economists tell us, such are the characteristics of a strongly competitive open commodity market. That's quite a story that lurks behind a simple chart of the size of the BGP routing table.
